WHAT IS CLAIMED IS:

1. A method of determining the biochemical or biophysical properties of a protein, said method comprising the steps of:

a) providing a database comprising protein sequence information and protein biochemical and/or biophysical properties,

b) analyzing the database using a data-mining technique,

c) correlating protein sequence, biochemical properties or biophysical properties, and

d) analyzing the sequence of the protein using the correlations to determine its biochemical or biophysical properties.

2. The method of claim 1, wherein the property being determined is a biophysical property.

3. The method of claim 2, wherein the biophysical property is thermal stability, solubility, isoelectric point, pH stability, crystalizability, conditions of crystallization, aggregation state, heat capacity (ΔC_p), resistance to chemical denaturation, resistance to proteolytic degradation, amide hydrogen exchange data, behavior on chromatographic matrices, electrophoretic mobility, resistance to degradation during mass spectrometry, and results obtained from nuclear magnetic resonance, X-ray crystallography, circular dichroism, light scattering, atomic adsorption, fluorescence, fluorescence quenching, mass spectroscopy, infrared spectroscopy, electron microscopy and atomic force microscopy.

4. The method of claim 3, wherein the biophysical property is thermal stability.

5. The method of claim 3, wherein the biophysical property is solubility.

6. The method of claim 3, wherein the biophysical property is crystalizability.

7. The method of claim 3, wherein the biophysical property is conditions of crystallization.

8. The method of claim 3, wherein the biophysical property is isoelectric point.

9. The method of claim 3, wherein the biophysical property is pH stability.

10. The method of claim 3, wherein the biophysical property is aggregation state.

11. The method of claim 3, wherein the biophysical property is heat capacity (ΔC_p).

12. The method of claim 3, wherein the biophysical property is resistance to chemical denaturation.

13. The method of claim 3, wherein the biophysical property is resistance to proteolytic degradation.

14. The method of claim 3, wherein the biophysical property is amide hydrogen exchange data.

15. The method of claim 3, wherein the biophysical property is behavior on chromatographic matrices.

16. The method of claim 3, wherein the biophysical property is electrophoretic mobility.

17. The method of claim 3, wherein the biophysical property is resistance to degradation during mass spectrometry.

18. The method of claim 3, wherein the biophysical property is the results obtained from nuclear magnetic resonance.

19. The method of claim 3, wherein the biophysical property is the results obtained from X-ray crystallography.

20. The method of claim 3, wherein the biophysical property is the results obtained from

circular dichroism.

21. The method of claim 3, wherein the biophysical property is the results obtained from light scattering.

22. The method of claim 3, wherein the biophysical property is the results obtained from atomic adsorption.

23. The method of claim 3, wherein biophysical property is the results obtained from fluorescence.

24. The method of claim 3, wherein the biophysical property is the results obtained from fluorescence quenching.

25. The method of claim 3, wherein the biophysical property is the results obtained from mass spectroscopy.

26. The method of claim 3, wherein the biophysical property is the results obtained from infrared spectroscopy.

27. The method of claim 3, wherein biophysical property is the results obtained from electron microscopy.

28. The method of claim 3, wherein the biophysical property is the results obtained from atomic force microscopy.

29. The method of claim 1, wherein the property being determined is a biochemical property.

30. The method of claim 29, wherein the biochemical property is expressability, protein yield, small-molecule binding, subcellular localization, utility as a drug target, protein-protein interactions or protein-ligand interactions.

31. The method of claim 30, wherein the biochemical property is small-molecule binding.

32. The method of claim 30, wherein the biochemical property is protein yield.

33. The method of claim 30, wherein the biochemical property is expressability.

34. The method of claim 30, wherein the biochemical property is subcellular localization.

35. The method of claim 30, wherein the biochemical property is utility as a drug target.

36. The method of claim 30, wherein the biochemical property is protein-protein interactions.

37. The method of claim 30, wherein the biochemical property is protein-ligand interactions.

38. The method of claim 1, wherein the data-mining technique is selected from the group decision-tree analysis, case-based reasoning, Bayesian classifier, simple linear discriminant analysis, and support vector machines.

39. The method of claim 38, wherein the data-mining technique is decision-tree analysis.

40. The method of claim 38, wherein the data-mining technique is case-based reasoning.

41. The method of claim 38, wherein the data-mining technique is Bayesian classifier.

42. The method of claim 38, wherein the data-mining technique is simple linear discriminant analysis.

43. The method of claim 38, wherein the data-mining technique is support vector machines.

44. A method of optimizing high-throughput protein structure determination, said method

comprising the steps of :

a) providing a database comprising protein sequence information and protein biochemical and biophysical properties,

b) analyzing the database using a data-mining technique,

c) determining correlations between protein sequence and biochemical or biophysical properties,

d) analyzing the sequence of a protein using said correlations to determine its biochemical or biophysical properties, and

e) optimizing the throughput of the protein structure determination based on said biochemical or biophysical properties by modifying the experimental procedures and/or modifying the protein sequence.

45. The method of claim 44, wherein the data-mining technique is selected from the group decision-tree analysis, case-based reasoning, Bayesian classifier, simple linear discriminant analysis, and

46. The method of claim 45, wherein the data-mining technique is decision-tree analysis.

47. The method of claim 45, wherein the data-mining technique is case-based reasoning.

48. The method of claim 45, wherein the data-mining technique is Bayesian classifier.

49. The method of claim 45, wherein the data-mining technique is simple linear discriminant analysis.

50. The method of claim 45, wherein the data-mining technique is support vector machines.

51. A method of optimizing high-throughput protein purification, said method comprising the steps of :

a) providing a database comprising protein sequence information and protein biochemical and biophysical properties,

b) analyzing the database using a data-mining technique,

c) determining correlations between protein sequence and biochemical or biophysical properties,

d) analyzing the sequence of a protein using the correlations to determine its biochemical or biophysical properties, and

e) optimizing the throughput of the protein purification based on said biochemical or biophysical properties by modifying the experimental procedures and/or modifying the protein sequence.

52. The method of claim 51, wherein the data-mining technique is selected from the group decision-tree analysis, case-based reasoning, Bayesian classifier, simple linear discriminant analysis, and support vector machines.

53. The method of claim 52, wherein the data-mining technique is decision-tree analysis.

54. The method of claim 52, wherein the data-mining technique is case-based reasoning.

55. The method of claim 52, wherein the data-mining technique is Bayesian classifier.

56. The method of claim 52, wherein the data-mining technique is simple linear discriminant analysis.

57. The method of claim 52, wherein the data-mining technique is support vector machines.

58. A method of optimizing high-throughput protein expression, said method comprising the steps of :

a) providing a database comprising protein sequence information and protein biochemical and biophysical properties,

b) analyzing the database using a data-mining technique,

c) determining correlations between protein sequence and biochemical or biophysical properties,

SUBA1)

d) analyzing the sequence of the protein using the correlations to determine its biochemical or biophysical properties, and

e) optimizing throughput of the protein expression based on said biochemical or biophysical properties by modifying the experimental procedures and/or modifying the protein sequence.

SUB A) Cont

59. The method of claim 58, wherein the data-mining technique is selected from the group decision-tree analysis, case-based reasoning, Bayesian classifier, simple linear discriminant analysis, and

60. The method of claim 59, wherein the data-mining technique is decision-tree analysis.

61. The method of claim 59, wherein the data-mining technique is case-based reasoning.

62. The method of claim 59, wherein the data-mining technique is Bayesian classifier.

63. The method of claim 59, wherein the data-mining technique is simple linear discriminant analysis.

64. The method of claim 59, wherein the data-mining technique is support vector machines.

65. A method of optimizing drug-target discovery, said method comprising the steps of :

a) providing a database comprising protein sequence information and protein biochemical and biophysical properties,

b) analyzing the database using a data-mining technique,

c) determining correlations between protein sequence and biochemical or biophysical properties,

d) analyzing the sequence of a protein using the correlations to determine its biochemical or biophysical properties, and

e) optimizing drug-target discovery base on said biochemical or biophysical properties by modifying the experimental procedures and/or modifying the protein

sequence.

66. A method of screening proteins for drug-target discovery, said method comprising the steps of :

a) providing a database comprising protein sequence information and protein biochemical and biophysical properties,

b) analyzing the database using a data-mining technique,

c) determining correlations between protein sequence and biochemical or biophysical properties,

d) analyzing the sequence of the protein using the correlations to determine its biochemical or biophysical properties, and

e) selecting proteins for analysis as a drug target based on their predicted biochemical and/or biophysical properties.